AN EXPLICIT BOUND ON THE LOGARITHMIC SOBOLEV CONSTANT OF WEAKLY DEPENDENT RANDOM VARIABLES

Katalin Marton

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences

Abstract. We prove logarithmic Sobolev inequality for measures

$$q^n(x^n) = \operatorname{dist}(X^n) = \exp\bigl(-V(x^n)\bigr), \quad x^n \in \mathbb{R}^n,$$

under the assumptions that:
(i) the conditional distributions

$$Q_i(\cdot|x_j, j \neq i) = \operatorname{dist}(X_i|X_j = x_j, j \neq i)$$

satisfy a logarithmic Sobolev inequality with a common constant ρ, and
(ii) they also satisfy some condition expressing that the mixed partial derivatives of the Hamiltonian V are not too large relative to ρ.

Condition (ii) has the form that the norms of some matrices defined in terms of the mixed partial derivatives of V do not exceed $1/2 \cdot \rho \cdot (1-\delta)$. The logarithmic Sobolev constant of q^n can then be estimated from below by $1/2 \cdot \rho \cdot \delta$. This improves on earlier results by Th. Bodineau and B. Helffer, by giving an explicit bound, for the logarithmic Sobolev constant for q^n.

1. Introduction.

Let us consider the absolutely continuous probability measures on the n-dimensional Euclidean space $\mathbb{R}^n$. With some abuse of notation, we use the same letter to denote a probability measure and its density function.

If p^n and q^n are density functions on the same Euclidean space $\mathbb{R}^n$, then $D(p^n\|q^n)$ and $I(p^n\|q^n)$ will denote, respectively, the relative entropy (called also informational divergence) and the Fisher information of p^n with respect to q^n:
If $q^n(x^n) = 0 \implies p^n(x^n) = 0$ then

$$D(p^n\|q^n) = \int_{\mathbb{R}^n} \log \frac{p^n(x^n)}{q^n(x^n)} p^n(x^n) dx^n,$$

otherwise $D(p^n\|q^n) = \infty$; if $\log \frac{p^n(x)}{q^n(x)}$ is smooth a.e.. then

$$I(p^n\|q^n) = \int_{\mathbb{R}^n} \left| \nabla \log \frac{p^n(x^n)}{q^n(x^n)} \right|^2 p(x^n) dx^n.$$

Definition. The density function q^n satisfies a logarithmic Sobolev inequality with constant $\rho > 0$ if for any density function p^n on $\mathbb{R}^n$, with $\log(p^n/q^n)$ smooth,

$$D(p^n\|q^n) \le \frac{1}{2\rho} I(p^n\|q^n).$$

A logarithmic Sobolev inequality for q^n can be used to control the rate of convergence to equilibrium for the diffusion process with limit distribution q^n, and is equivalent to the hypercontractivity of the associated semigroup. The prototype is Gross' logarithmic Sobolev inequality for the Gaussian measures which is associated with the Ornstein-Uhlenbeck process [Gr]. Another use of logarithmic Sobolev inequalities is to derive transportation cost inequalities, a tool to prove measure concentration (F. Otto, C. Villani [O-V]). We shall return to the Otto-Villani theorem later in the introduction. The logarithmic Sobolev inequality for spin systems is equivalent to the property called "exponential decay of correlation"; for this concept we refer to Stroock and Zegarlinski, [SZ1], [SZ2], [SZ3], Zegarlinski [Z], Bodineau and Helffer [B-H] and Helffer [H].

Our aim is to prove a logarithmic Sobolev inequality for measures with positive density

$$q^n(x^n) = \text{dist}(X^n) = \exp\bigl(-V(x^n)\bigr), \quad x^n \in \mathbb{R}^n,$$

under the assumption that the conditional distributions

$$Q_i(\cdot|x_j, j \neq i) = \text{dist}(X_i|X_j = x_j, j \neq i)$$

satisfy a logarithmic Sobolev inequality with a constant ρ independent of i and x^n, and also some condition expressing that the mixed partial derivatives of V are not too large relative to ρ. We want a logarithmic Sobolev constant independent of n.

Much work has been done on this subject. When the range of the X_i's is finite then another definition is needed for Fisher information, and thus for logarithmic Sobolev inequality. See [D-H], [SZ2], [SZ3]. The case when the range of the X_i's is finite or compact was studied by J-D. Deuschel and D. W. Holley [D-H], D. Stroock and B. Zegarlinski [S-Z1], [S-Z2], [S-Z3], L. Lu and H. T. Yau [L-Y] and others. In [S-Z3] Stroock and Zegarlinski prove equivalence between logarithmic Sobolev inequality and Dobrushin and Shlosman's strong mixing condition in the discrete case. The discrete or compact cases are simpler, basically because of the following reason: A random variable with finite range automatically admits a logarithmic Sobolev inequality with a constant depending only on the size of the range. The same holds for a one-dimensional compact range, but not for a one-dimensional non-compact range.

The non-compact case was studied by N. Yoshida [Y1], [Y2], B. Helffer [He], B. Helffer and Th. Bodineau [Bod-He] and M. Ledoux [L]. Their results assert the existence of a positive logarithmic Sobolev constant under (more or less) the above conditions. But these results do not say much about how small the mixed partial derivatives of V should be cmpared to the logarithmic Sobolev constant of the $Q_i(\cdot|\bar{x}_i)$'s, and do not provide an explicit lower bound on the logarithmic Sobolev constant of q^n. Our aim is to improve on earlier results in this respect.

We shall use the following

Notation:

- For $x^n = (x_1, x_2, \dots x_n) \in \mathbb{R}^n$ and $1 \le i \le n$,
$$\bar{x}_i = (x_j : j \ne i), \quad x^i = (x_j : j \le i), \quad x_i^n = (x_j : i < j \le n);$$
- $p^n :$ density of an absolutely continuous probability measure on $\mathbb{R}^n$;
- $Y^n = Y^n(1) :$ random sequence in $\mathbb{R}^n$, $\mathrm{dist}(Y^n) = p^n$;
- $p_i(\cdot|\bar{y}_i) = \mathrm{dist}(Y_i|\bar{Y}_i = \bar{y}_i) \quad (1 \le i \le n) :$ conditional density functions consistent with p^n;
- $\bar{p}_i := \mathrm{dist}(\bar{Y}_i), \quad p^i = \mathrm{dist}(Y^i), \quad p_i^n(\cdot|y^i) = \mathrm{dist}(Y_i^n|Y^i = y^i)$.
- $q^n :$ density of a fixed absolutely continuous probability measure on $\mathbb{R}^n$;
- $X^n :$ random sequence in $\mathbb{R}^n$, $\mathrm{dist}(X^n) = q^n$;
- $\mathcal{Q} = \big(Q_i(x_i|\bar{x}_i), \quad 1 \le i \le n, \quad x_i \in \mathbb{R} \quad \bar{x}_i \in \mathbb{R}^{n-1}\big) :$ system of conditional density functions; usually $Q_i(\cdot|\bar{x}_i) = \mathrm{dist}(X_i|\bar{X}_i = \bar{x}_i) \quad (1 \le i \le n)$.

The following sufficient condition for logarithmic Sobolev inequalities is well known; it follows from the Bakry and Emery's celebrated criterion [Ba-E], supplemented by a perturbation result by Holley and Stroock [Ho-S]:

Theorem 0. *Let $q(x^n) = \exp(-V(x^n))$ be a density function on $\mathbb{R}^k$, and let V be strictly convex at ∞, i.e., $V(x^k) = U(x^k) + K(x^k)$, where*

$$\mathit{Hess}(U)(x^k) = \big(\partial_{ij} U(x^k)\big) \geq c \cdot I_n$$

for some $c > 0$ (where I_n is the identity matrix), and $K(x^k)$ is bounded. Then q satisfies a logarithmic Sobolev inequality with constant ρ, depending on c and $\|K\|_\infty$:

$$\rho \geq c \cdot \exp\big(-4\|K\|_\infty\big).$$

In particular, if V is uniformly strictly convex, i.e., $\mathit{Hess}(V) \geq c \cdot I_n$, then $\rho \geq c$.

Note that S. Bobkov and F. Götze [B-G] established a characterization of density functions on $\mathbb{R}$, satisfying a logarithmic Sobolev inequality.

We mention also the very important fact that a product distribution admits a logarithmic Sobolev inequality with constant ρ, provided the factors have logarithmic Sobolev constants $\geq \rho$.

In particular, a product distribution where all factors are uniformly bounded perturbations of uniformly log-concave distributions, admits a logarithmic Sobolev inequality with a controllable constant. The simplest case beyond this is when all the $Q_i(\cdot|\bar{x}_i)$'s are uniformly bounded perturbations of uniformly log-concave distributions, but there is a weak dependence between the coordinates. This was the case investigated by N. Yoshida [Y1], [Y2], B. Helffer [He], B. Helffer and Th. Bodineau [Bod-He] and M. Ledoux [L], and this is the theme of the present paper.

We proceed to the notations and definitions needed to formulate the main result.

Let us be given a system of conditional density functions

$$\mathcal{Q} = \big(Q_i(x_i|\bar{x}_i), \quad 1 \leq i \leq n, \quad x_i \in \mathbb{R} \quad \bar{x}_i \in \mathbb{R}^{n-1}\big). \tag{1.1}$$

We consider $Q_i(x_i|\bar{x}_i)$ as a function of n variables $x_1, x_2, \ldots, x_n$, and ∂_i will denote partial derivative with respect to the i'th variable. Define the triangular function matrices B_1 and B_2 as follows: For $y^n, z^n, \eta^n \in \mathbb{R}^n$ put

$$\beta_{i,k}(y^n, \eta^n, z^n) = \partial_{i,k}\big(-\log Q_i(\eta_i|y^{i-1}, z_i^n)\big),$$

and define

$$B_1(y^n, \eta^n, z^n) = \big(\beta_{i,k}(y^n, \eta^n, z^n)\big)_{i<k}, \quad B_2(y^n, \eta^n, z^n) = \big(\beta_{i,k}(y^n, \eta^n, z^n)\big)_{i>k}.$$

If $\mathcal{Q}$ consists of the conditional density functions of $q^n(x^n) = \exp(-V(x^n))$ then

$$\beta_{i,k}(y^n, \eta^n, z^n) = \partial_{i,k} V(y^{i-1}, \eta_i, z_i^n),$$

since $-\log Q_i(x_i|\bar{x}_i)$ and $V(x^n)$ differ in an additive term not depending on x_i.

Definition. Let us assume that the conditional density functions (1.1) satisfy the logarithmic Sobolev inequality with constant ρ independent of i and $\bar{x}_i$. We say that the system $\mathcal{Q}$ in (1.1) satisfies the $1/2(1-\delta)$-contracivity condition for partial derivatives if

$$\sup_{y^n,\eta^n,z^n}\left\|\frac{1}{\rho}B_j(y^n,\eta^n,z^n)\right\|\le\frac{1}{2}\cdot(1-\delta),\quad j=1,2. \tag{C}$$

We shall use the following short-hand

Notation.
A system $\mathcal{Q}$ of conditional density functions satisfies condition $LSI\&C(\rho,1/2(1-\delta))$ if the $Q_i(\cdot|\bar{x}_i)$'s satisfy a logarithmic Sobolev inequality with constant ρ, and also the $1/2(1-\delta)$-contracivity condition for partial derivatives.

The aim of this paper is to prove the following

Theorem.
If the conditional density functions $Q_i(\cdot|\bar{x}_i)$ satisfy condition $LSI\&C(\rho,1/2(1-\delta))$ then there exists a unique density function $q^n(x^n)$ on $\mathbb{R}^n$ such that its conditional density functions are the functions (1.1), and q^n satisfies the logarithmic Sobolev inequality

$$D(p^n\|q^n)\le\frac{1}{\rho\delta}I(p^n\|q^n). \tag{1.2}$$

I.e., the logarithmic Sobolev constant of q^n is at least $\rho\delta/2$.

This is clearly a perturbation result.

Inequalty (1.2) can also be written in the form

$$D(p^n\|q^n)\le\frac{1}{\rho-2\cdot\sup\max\{||B_1||,||B_2||\}}I(p^n\|q^n). \tag{1.3}$$

Note that the $LSI\&C(\rho,1/2(1-\delta))$ condition depends on the system of coordinates. This is a serious drawback, although in the case of spin systems this is natural. Moreover, the $LSI\&C(\rho,1/2(1-\delta))$ condition also depends on the ordering of coordinates (see later).

Because of the dependence on the system of coordinates, there are important families of distributions $q^n(x^n)=\exp(-V(x^n))$ (with n growing) that admit a logarithmic Sobolev inequality with a constant independent of n, without satisfying the conditions of our Theorem. In fact, this is the case with many convex quadratic functions $V(x^n)$, e.g.,

$$V_0(x^n)=1/2\cdot\sum_{i=1}^n x_i^2+1/2\cdot\big(\sum_{i=1}^n x_i-M\big)^2+\text{const.},\qquad M\in\mathbb{R}\quad\text{fixed.}$$

In a paper on conservative spin systems Landim, Panizo and Yau [L-P-Y] proved logarithmic Sobolev inequality for the following class of densities $\exp(-V(x^n))$:

$$V(x^n) = V_0(x^n) + \sum_{i=1}^{n} \phi(x_i),$$

where $\phi : \mathbb{R} \mapsto \mathbb{R}$ is bounded, and has bounded first and second derivatives. It would be nice to find a common way to prove our perturbative theorem and the theorem of Landim, Panizo and Yau [L-P-Y], but so far these two directions could not be united, and in fact [L-P-Y] contains the only non-perturbative result for the non-compact case.

The definition of the $1/2(1-\delta)$-contracivity condition is not very transparent, partly because it does depend on the ordering of the index set. If the indices are nods in a lattice in a Euclidean space, and if V is sufficiently symmetric, then the following consideration may help. The definition of V can often be extended in a natural way to infinite sequences $y = (y_i)$ indexed by the nods of the entire lattice. Let us consider the lexicographical ordering on the nods (i.e., on the index set). For every nod i, the symmetry with center i is a bijection between the nods precedeing resp. following i, and it often happens that if nods j and k are interchanged by this bijection then

$$\beta_{i,j}(y, \eta, z) = \beta_{i,k}(z^*, \eta, y^*),$$

where y^* denotes the sequence defined as follows:

$$y_i^* = y_i, \quad y_k^* = y_j \quad \text{if} \quad j \quad \text{and} \quad k \quad \text{are interchanged by the bijection.}$$

With this ordering of the indices it is often possible to give effective bounds for $\|B_1\|$ and $\|B_2\|$. Moreover, for the important special case when the mixed partial derivatives are constants $\beta_{i,k}$, the matrices B_1, B_2 have numerical entries. Also, we can use the following estimates: Denoting $\alpha_{i,k} = \sup_{x^n} |\partial_{i,k} V(x^n)|$, and writing

$$A_1 = (\alpha_{i,k})_{i<k}, \quad A_2 = (\alpha_{i,k})_{i>k},$$

we have

$$\|B_j\| \le \|A_j\|, \quad j = 1, 2.$$

One of the basic tools in the proof of the Theorem is a result by Otto and Villani [O-V], establishing a connection between the logarithmic Sobolev inequality and a transportation cost inequality for quadratic Wasserstein distance. (A transportation cost inequality bounds some distance between measures by the relative entropy of the measures.)

Definition. The (quadratic) Wasserstein distance, or W-distance, between two probability measures p^n and and q^n (in $\mathbb{R}^n$) is:

$$W(p^n, q^n) = \inf_{\pi} [E_\pi |Y^n - X^n|^2]^{1/2},$$

where Y^n and X^n are random variables with laws p^n resp. q^n, and infimum is taken over all distributions π on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals p^n and q^n.

A transportation cost inequality for quadratic Wasserstein distance was first proved by M. Talagrand [T], in the case of Gaussian distributions. It was generalized by Otto and Villani, who revealed a deep connection between logarithmic Sobolev and transportation cost inequalities.

Theorem of Otto and Villani. *[O-V], [B-G-L] If the density function $q^n(x^n)$ satisfies a logarithmic Sobolev inequality with constant ρ then*

$$W^2(p^n, q^n) \le \frac{2}{\rho} \cdot D(p^n \| q^n).$$

This form of the theorem was proved in [B-G-L]; the original formulation in [O-V] contained some minor additional condition. We shall use this theorem in dimension 1, except for the proof of the Corollary below.

It is not known whether a transportation cost inequality for W-distance does imply a logarithmic Sobolev inequality; the conjecture is no.

The importance of the Otto-Villani theorem for us lies in the following: By the $LSI\&C(\rho, 1/2(1-\delta))$ condition we can, roughly speaking, estimate $D(Q_i(\cdot|\bar{x}_i) \| Q_i(\cdot|\bar{z}_i))$ in terms of the Euclidean distance of the sequences $\bar{x}_i$ and $\bar{z}_i$. By the Otto-Villani theorem this translates to a bound for $W^2(Q_i(\cdot|\bar{x}_i), Q_i(\cdot|\bar{z}_i))$. Thereby we derive from the $LSI\&C(\rho, 1/2(1-\delta))$ condition an analog of Dobrushin's uniqueness condition [D]. This will assure a contractivity property for a Markov chain that we construct to interpolate between p^n and q^n. (See Section 5, Lemma 3). Such an interpolation was also used in [M2].

Related to such estimates is the following corollary which may be interesting on its own right:

Corollary to the Theorem. *If the conditional density functions of q^n satisfy condition $LSI\&C(\rho, 1/2(1-\delta))$ then for any $i \in [1, n]$, and any sequences x^i, $\hat{x}^i$*

$$W^2\big(q_i^n(\cdot|x^i), q_i^n(\cdot|\hat{x}^i)\big) \le \frac{4}{\rho\delta} \cdot D(q_i^n(\cdot|x^i) \| q_i^n(\cdot|\hat{x}^i)) \le \left(\frac{1-\delta}{\delta}\right)^2 \cdot |x^i - \hat{x}^i|^2. \quad (1.4)$$

I.e., the mapping $x^i \mapsto q_i^n(\cdot|x^i)$ is Lipschitz with respect to the Euclidean and quadratic Wasserstein distances, with Lipschitz constant $(1-\delta)/\delta$.

The proof of the Corollary (Section 5) will make it clear that for (1.4) we only need the Theorem for $n-1$. This allows to use the Corollary in a proof by induction.

In an earlier paper [M2, Theorem 2] we considered distributions q^n satisfying conditions similar to $LSI\&C(\rho, 1/2(1-\delta))$, and proved a transportation cost inequality. In view of the Otto-Villani theorem, this is weaker than a logarithmic Sobolev inequality. The contractivity condition for partial derivatives in the present paper is somewhat stronger than the one in Theorem 2 of [M2]. Note that in [M2] we also considered the more general case dealing with conditional density functions $Q_I(\cdot|\bar{x}_I)$, where I runs over a collection of (small) subsets of $[1, n]$, and $Q_I(\cdot|\bar{x}_I)$ is the joint conditional density function of the random variables $(X_i : i \in I)$, given the values $(X_j : j \notin I)$. We did not try to achieve this generality in the present paper.

The proof of the Theorem is quite different from the approach taken by Bodineau and Helffer, who derived logarithmic Sobolev inequality from the exponential decay of correlations. Our proof is based on a discrete time interpolation connecting the distributions p^n and q^n. This interpolation is more complicated than the Gibbs sampler used in [M2], and may seem somewhat artificial. Moreover, it does not seem applicable for stochastic simulation. However, we could not find any simpler interpolation doing the job.

Even apart from the restriction of generality, the present paper does not make Theorem 2 of [M2] superfluous. The proof of Theorem 2 in [M2], at least for the case considered in the present paper, is much simpler, and has more intrinsicly understandable meaning. Moreover, Theorem 1 in [M2] is not implied by the present result.

2. An auxiliary theorem for estimating relative entropy.

We shall use the following notation: If Y and X are random variables with values in the same space, and with distributions p and q, respectively, then we denote by $D(Y\|X)$ the relative entropy $D(p\|q)$. Moreover, if we are given the conditional distributions $\mathrm{dist}(Y|V)$, $\mathrm{dist}(X|U)$ and the joint distribution $\pi = \mathrm{dist}(U, V)$, then we denote by $D(Y|U\|X|V)$ the following average relative entropy:

$$D(Y|V\|X|U) = \mathbb{E}_\pi D(Y|V = v\|X|U = u),$$

where $\mathbb{E}_\pi$ denotes expectation with respect to π. Also, we use the notation

$$D(Y|V\|P(\cdot|U)) = \mathbb{E}_\pi D(Y|V = v\|P(\cdot|U = u)),$$

where $P(\cdot|U)$ denotes a conditional distribution.

In this section we prove the following

Auxiliary Theorem. *Let X^n be a random sequence with conditional density functions $Q_i(\cdot|\bar{x}_i)$, and let $\big(Y^n(t) : t = 1, 2, \dots\big)$ be a discrete time random process in $\mathbb{R}^n$. Then for any $s \geq 1$*

$$\begin{aligned} &D(Y^n(1)\|X^n) \\ &\leq \sum_{t=1}^{s}\sum_{i=1}^{n} D\big(Y_i(t)|Y^{i-1}(t), Y_i^n(t+1)\|Q_i(\cdot|Y^{i-1}(t), Y_i^n(t+1)\big) \\ &+ D(Y^n(s+1)\|X^n). \end{aligned} \tag{2.1}$$

In particular, if

$$\lim_{s\to\infty} D(Y^n(s)\|X^n) = 0 \quad \textit{along some subsequence} \tag{2.2}$$

then

$$\begin{aligned} &D(Y^n(1)\|X^n) \\ &\leq \sum_{t=1}^{\infty}\sum_{i=1}^{n} D\big(Y_i(t)|Y^{i-1}(t), Y_i^n(t+1)\|Q_i(\cdot|Y^{i-1}(t), Y_i^n(t+1))\big). \end{aligned} \tag{2.3}$$

Remark.
A frequently used tool in bounding relative entropy is the decomposition

$$D(Y^n\|X^n) = \sum_{i=1}^{n} D(Y_i|Y^{i-1}\|q_i(\cdot|Y^{i-1})),$$

where $q_i(\cdot|y^{i-1}) = \mathrm{dist}(X_i|X^{i-1} = y^{i-1})$. This decomposition has the drawback that $D(Y_i|Y^{i-1}\|q_i(\cdot|Y^{i-1})$ cannot be easily bounded by a relative entropy with respect to $Q_i(\cdot|\bar{Y}_i)$ or anything similar. The Auxiliary Theorem bounds $D(Y^n\|X^n)$ by an infinite sum of relative entropies, all with respect to some conditional density function of form $Q_i(\cdot|\cdot)$.

Proof.
First we prove (2.1) for $s = 1$:

$$\begin{aligned} &D(Y^n(1)\|X^n) \\ &\leq \sum_{i=1}^{n} D\big(Y_i(1)|Y^{i-1}(1), Y_i^n(2)\|Q_i(\cdot|Y^{i-1}(1), Y_i^n(2)\big) + D(Y^n(2)\|X^n). \end{aligned} \tag{2.4}$$

To this end, we prove the following by recursion on i:

$$\begin{aligned} D(Y^n(1)\|X^n) &\leq \sum_{j=i}^{n} D(Y_j(1)|Y^{j-1}(1), Y_j^n(2)\|Q_j(\cdot|Y^{j-1}(1), Y_j^n(2))) \\ &+ D\big((Y^{i-1}(1), Y_{i-1}^n(2))\|X^n\big), \quad i = 1, 2, \dots, n. \end{aligned} \tag{2.5}$$

For $i = 1$ this is just (2.4).

To prove (2.5) for $i = n$, we use the expansion of relative entropy:

$$\begin{aligned} D(Y^n(1)\|X^n) &= D(Y^{n-1}(1)\|X^{n-1}) + D(Y_n(1)|Y^{n-1}(1)\|Q_n(\cdot|Y^{n-1}(1))) \\ &\leq D(Y_n(1)|Y^{n-1}(1)\|Q_n(\cdot|Y^{n-1}(1))) + D\big((Y^{n-1}(1), Y_n(2))\|X^n\big). \end{aligned} \tag{2.6}$$

Applying (2.6) for $D(Y^{n-1}(1), Y_n(2)\|X^n)$, we get

$$\begin{aligned} &D\big(Y^{n-1}(1), Y_n(2)\|X^n\big) \\ &\leq D(Y_{n-1}(1)|Y^{n-2}(1), Y_n(2)\|Q_{n-1}(\cdot|Y^{n-2}(1), Y_n(2))\big) \\ &+ D\big(Y^{n-2}(1), Y_{n-1}(2), Y_n(2)\|X^n\big). \end{aligned}$$

Substituting this into (2.6), we get (2.5) for $i = n - 1$:

$$\begin{aligned} &D(Y^n(1)\|X^n) \leq D(Y_n(1)|Y^{n-1}(1)\|Q_n(\cdot|Y^{n-1}(1))) \\ &+ D\big(Y_{n-1}(1)|Y^{n-2}(1), Y_n(2)\|Q_{n-1}(\cdot|Y^{n-2}(1), Y_n(2))\big) \\ &+ D\big(Y^{n-2}(1), Y_{n-1}(2), Y_n(2)\|X^n\big). \end{aligned}$$

This procedure can be continued to get (2.5) for every i, thus proving (2.4).

Applying (2.4) to $D(Y^n(2)\|X^n)$, we get

$$\begin{aligned} &D(Y^n(2)\|X^n) \\ &\leq \sum_{i=1}^{n} D\big(Y_i(2)|Y^{i-1}(2), Y_i^n(3)\|Q_i(\cdot|Y^{i-1}(2), Y_i^n(3))\big) + D(Y^n(3)\|X^n). \end{aligned}$$

Substituting this into (2.4), we get (2.1) for $s = 2$. It is clear that the same argument applied to $D(Y^n(t)\|X^n)$, will prove (2.1) for any s. $\square$

The proof of the Auxiliary Theorem is fairly simple, but to use it we need a process $(Y^n(t), t = 1, 2, \ldots)$ that admits good estimates for the terms in the sum (2.3), and also satisfies (2.2). The construction and analysis of such a process is the subject of the rest of the paper.

3. Preliminaries for the construction.

We shall use the following

Notation. Let (T, U, V) be a triple of random variables. We write

$$T \to U \to V$$

to express the Markov relation

$$T \quad \text{and} \quad V \quad \text{are conditionally independent given} \quad U.$$

The following definition and lemma are crucial in the forthcoming construction.

Definition. Consider a pair of random sequences (Y^n, S^n) in $\mathbb{R}^n \times \mathbb{R}^n$. We say that (Y^n, S^n) (or its distribution) is loosely connected if for every $i \in [1, n]$

$$Y_i \to (Y^{i-1}, S_i^n) \to S^i. \tag{3.1}$$

Lemma 1.
(i) For any joint distribution dist(Y^n, S^n), there exists a dist$(\hat{Y}^n, \hat{S}^n)$ such that, for every $i \in [1, n]$, we have dist$(\hat{Y}^i, \hat{S}_i^n)$ = dist(Y^i, S_i^n), and $(\hat{Y}^n, \hat{S}^n)$ is loosely connected.
(ii) The joint distribution (Y^n, S^n) is loosely connected if and only if (S^n, Y^n), with the ordering of the indices reversed, is such, i.e., iff

$$S_i \to (Y^{i-1}, S_i^n) \to Y_{i-1}^n, \quad 1 \le i \le n. \tag{3.2}$$

Proof.
(i): Define the joint distribution $\mathrm{dist}(\hat{Y}_1, \hat{S}^n)$ by

$$\mathrm{dist}(\hat{S}^n) = \mathrm{dist}(S^n), \quad \mathrm{dist}(\hat{Y}_1, \hat{S}_1^n) = \mathrm{dist}(Y_1, S_1^n) \quad \text{and} \quad \hat{Y}_1 \to \hat{S}_1^n \to \hat{S}_1.$$

Let us assume that we have already defined $\mathrm{dist}(\hat{Y}^{i-1}, \hat{S}^n)$ in such a way that $\mathrm{dist}(\hat{S}^n) = \mathrm{dist}(S^n)$, and for $k \le i-1$

$$\mathrm{dist}(\hat{Y}^k, \hat{S}_k^n) = \mathrm{dist}(Y^k, S_k^n) \quad \text{and} \quad \hat{Y}_k \to (\hat{Y}^{k-1}, \hat{S}_k^n) \to \hat{S}^k. \tag{3.3}$$

We can extend this definition to a joint distribution $\mathrm{dist}(\hat{Y}^i, \hat{S}^n)$ so that (3.3) is preserved for $k \le i$. Indeed, we define $\mathrm{dist}(\hat{Y}^i, \hat{S}_i^n) = \mathrm{dist}(Y^i, S_i^n)$, which is possible since $\mathrm{dist}(\hat{Y}^{i-1}, \hat{S}_i^n) = \mathrm{dist}(Y^{i-1}, S_i^n)$, and then make

$$\hat{Y}_i \to (\hat{Y}^{i-1}, \hat{S}_i^n) \to \hat{S}^i.$$

So the proof can be completed by recursion.

(ii): It is clear that (3.1) implies

$$S_i \to (Y^{i-1}, S_i^n) \to Y_i. \tag{3.4}$$

Moreover, applying (3.1) for $i+1$,

$$(S_i, S_{i+1}) \to (Y^i, S_{i+1}^n) \to Y_{i+1},$$

whence

$$S_i \to (Y^i, S_i^n) \to Y_{i+1}. \tag{3.5}$$

Putting together (3.4) and (3.5), we see that

$$S_i \to (Y^{i-1}, S_i^n) \to (Y_i, Y_{i+1}).$$

Iterating this reasoning we arrive at (3.2). □

Definition. The random pair $(\hat{Y}^n, \hat{S}^n)$ in Lemma 1 will be called the loosely connected copy of (Y^n, S^n).

A basic step in the construction of the process $(Y^n(t))$, to be used in the Auxiliary Theorem, is singled out in the next definition.

Definition. Let us be given the system of conditional distributions $\mathcal{Q}$ as in (1.1), and a loosely connected joint distribution $\operatorname{dist}(Y^n, S^n)$. We extend $\operatorname{dist}(Y^n, S^n)$ to

$$\operatorname{dist}(Y^n, S^n, \eta^n) \tag{3.6}$$

as follows. First we define, for every i, the conditional distribution $\operatorname{dist}(\eta_i | Y^n, S^n)$. To this end, put

$$\operatorname{dist}\big(\eta_i | Y^{i-1}, S_i^n\big) = Q_i(\cdot | Y^{i-1}, S_i^n).$$

Then define

$$\operatorname{dist}(Y_i, \eta_i | Y^{i-1}, S_i^n) \tag{3.7}$$

as that joining of the conditional distributions

$$\operatorname{dist}(Y_i | Y^{i-1}, S_i^n) \qquad \text{and} \qquad Q_i(\cdot | Y^{i-1}, S_i^n), \tag{3.8}$$

that achieves W-distance for any given value of the conditions. Thereby we have defined $\operatorname{dist}(\eta_i | Y^i, S_i^n)$. Then put

$$\operatorname{dist}(\eta_i | Y^n, S^n) = \operatorname{dist}(\eta_i | Y^i, S_i^n),$$

and make the random variables η_i, $i = 1, \ldots, n$, conditionally independent of each other, given (Y^n, S^n).

The random triple (3.6) will be called the $\mathcal{Q}$-extension of $\operatorname{dist}(Y^n, S^n)$.

4. The interpolation process and the Main Lemma.

Let us be given a loosely connected joint distribution

$$\operatorname{dist}(Y^n(0), Y^n(1)), \tag{4.1}$$

and the system of conditional density functions $\mathcal{Q}$ (c.f. (1.1)). Starting from the distribution (4.1), we are going to construct a discrete time process

$$\big(Y^n(t) : t = 0, 1, 2, \dots\big) \tag{4.2}$$

for which then the Auxiliary Theorem will be applied to estimate $D(Y^n(1)\|X^n)$. The evolution of the process will be governed, in a tricky way, by the conditional density functions $Q_i(\cdot|\bar{x}_i)$. The process (4.2) can be considered as an interpolation between $Y^n(1)$ and X^n.

Actually we are going to build a process

$$\big((Y^n(t-1), \eta^n(t+1)) : t = 1, 2, \dots\big). \tag{4.3}$$

We want this process to satisfy the following requirements:

(i) The sequence $(Y^n(t), t = 0, 1, 2, \dots)$ is a Markov chain;

(ii) the $\eta^n(t)$'s are conditionally independent of each other, given the entire process $\big(Y^n(t), t = 0, 1, \dots\big)$;

(iii) for $t \geq 1$,

$$\eta^n(t+1) \to (Y^n(t-1), Y^n(t)) \to \big(Y^n(s), s \neq t-1, s \neq t\big);$$

(iv) For $t \geq 1$, $(Y^n(t-1), Y^n(t), \eta^n(t+1))$ is the $\mathcal{Q}$-extension of $(Y^n(t-1), Y^n(t))$;

(v) for $t \geq 2$ and all i

$$\begin{aligned}&\operatorname{dist}(Y_i(t+1)|Y_i^n(t+1), Y^n(s), s \leq t) = \operatorname{dist}(Y_i(t+1)|Y^{i-1}(t), Y_i^n(t+1))\\ &= \operatorname{dist}(\eta_i(t+1)|\eta^{i-1}(t), \eta_i^n(t+1)).\end{aligned} \tag{4.4}$$

For $t = 1$ we require only that $(Y^n(1), Y^n(2))$ be loosely connected, and that $\operatorname{dist}(Y^n(2)) = \operatorname{dist}(\eta^n(2))$.

Requirement (v) implies that $\operatorname{dist}(Y^n(t), Y^n(t+1))$ is the loosely connected copy of $\operatorname{dist}(\eta^n(t), \eta^n(t+1))$ (see (ii) of Lemma 1), in particular $\operatorname{dist}(Y^n(t)) = \operatorname{dist}(\eta^n(t))$ for $t \geq 2$.

We construct the process (4.3) by induction on t. Let us assume that we have already defined

$$\operatorname{dist}((Y^n(0), Y^n(s), \eta^n(s+1)) : 1 \leq s \leq t) \tag{4.5}$$

with properties (i-v). (For $t = 1$ this obviously can be done.) We extend (4.5) to

$$\mathrm{dist}((Y^n(0), Y^n(s), \eta^n(s+1)) : 1 \le s \le t+1). \tag{4.6}$$

To this end, first we define

$$\mathrm{dist}(Y^n(s), s \le t+1) \tag{4.7}$$

so as to satisfy (4.4). (We use the fact that $\mathrm{dist}(Y^n(t)) = \mathrm{dist}(\eta^n(t))$, implied by the induction hypothesis, and define recursively, for $i = n, n-1, \dots, 1$,

$$\mathrm{dist}(Y_i(t+1)|Y_i^n(t+1), Y^n(s), s \le t).)$$

Having defined (4.7), we make $Y^n(t+1)$ conditionally independent of $(\eta^n(s), s \le t+1)$, given $(Y^n(s), s \le t)$. Then we complete the definition of (4.6), extending (4.7) by $\eta^n(t+2)$, so as to satisfy (ii), (iii) and (iv). The inductive construction is complete.

Note that the joint distribution (4.1) and the conditional density functions (1.1) do not completely determine the process $(Y^n(t))$, since condition (v) allows some freedom in the definition of $\mathrm{dist}(Y^n(1), Y^n(2))$.

In the rest of the paper $\big(Y^n(t)\big) = \big(Y^n(t) : t = 0, 1, 2, \dots\big)$ will always denote the process (4.2), i.e., a process built up as in this section, starting from a joint distribution (4.1), governed by the conditional density functions (1.1), and satisfying properties (i-v).

We are going to use the Auxiliary Theorem for the process (4.2). First we formulate an intermediate result. We shall use the following notation for the terms on the right-hand-side of (2.3): For $t \ge 0$ we write

$$D_t = \sum_{i=1}^{n} D\big(Y_i(t)|Y^{i-1}(t), Y_i^n(t+1) \| Q_i(\cdot|Y^{i-1}(t), Y_i^n(t+1))\big). \tag{4.8}$$

Thus the sum on the right-hand-side of (2.3) is equal to $\sum_{t=1}^{\infty} D_t$. (Note that D_0 is not included in the sum.)

In Sections 5-7. we prove the following

Main Lemma. *If the system of conditional density functions $\mathcal{Q}$ satisfies condition $LSI\&C(\rho, 1/2(1-\delta))$ then there exists a random sequence X^n such that the conditional density functions of X^n are the functions Q_i, and*

$$D(Y^n(1)\|X^n) \le \frac{1/2(1-\delta)^2 \cdot D_0 + D_1}{1-(1-\delta)^2}. \tag{4.9}$$

The proof of the Main Lemma articulates as follows: After some preliminaries, in Section 6 we show that the sum in (2.3) is overbounded by the right-hand-side of (4.9). Then in Section 7 we prove the convergence (2.2). Finally, In Section 8 we estimate D_0 and D_1 in terms of $I(p^n\|q^n)$, where $p^n = \mathrm{dist}(Y^n(1))$.

5. Preliminaries for the proof of the Main Lemma.

We shall need the following lemma:

Lemma 2.
For any joint distributions dist(U, V)*, dist*(U, Y) *and dist*(V, X)*,*

$$D(Y\|X) \leq D(Y|U\|X|V).$$

Proof. Write $d\pi(u, v) = \mathrm{dist}(U, V)$. Define $\mathrm{dist}(Y|U = u, V = v) = \mathrm{dist}(Y|U = u)$, and $\mathrm{dist}(X|U = u, V = v) = \mathrm{dist}(X|V = v)$. Then

$$\mathrm{dist}(Y) = \int \mathrm{dist}(Y|U = u, V = v) d\pi(u, v),$$

and

$$\mathrm{dist}(X) = \int \mathrm{dist}(X|U = u, V = v) d\pi(u, v).$$

It is known that $D(r\|s)$ (r, s probability distributions) is a convex function of the pair (r, s). Therefore, the above representations imply

$$D(Y\|X) \leq \int D\big(Y|U = u, V = v\|X|U = u, V = v\big) d\pi(u, v) = D(Y|U\|X|V). \quad \square$$

The analysis of the process (4.2) is based on Lemmas 3 and 4 below.

Lemma 3.
Assume that the conditional density functions $Q_i(\cdot|\bar{x}_i)$ *satisfy condition* $LSI\&C(\rho, 1/2(1 - \delta))$*. Then:*

(i) For any quintuple of sequences $\big(u^n(1), u^n(2), t^n(1), t^n(2), \zeta^n\big)$*:*

$$\sum_{i=1}^{n} \left| \partial_i \log \frac{Q_i(\zeta_i|u^{i-1}(1), u_i^n(2))}{Q_i(\zeta_i|t^{i-1}(1), t_i^n(2))} \right|^2 \leq 1/2 \cdot \rho^2 \cdot (1 - \delta)^2 \cdot \Big[|u^n(1) - t^n(1)|^2 + |u^n(2) - t^n(2)|^2 \Big]. \tag{5.1}$$

If $u^n(1) = t^n(1)$ *then the right-hand-side can be divided by* 2.
(ii) For any quadruple of random sequences $\big(U^n(1), U^n(2), T^n(1), T^n(2)\big)$*:*

$$\sum_{i=1}^{n} D\big(Q_i(\cdot|U^{i-1}(1), U_i^n(2)) \| Q_i(\cdot|T^{i-1}(1), T_i^n(2))\big) \leq 1/4 \cdot \rho \cdot (1 - \delta)^2 \cdot \mathbb{E}\Big[|U^n(1) - T^n(1)|^2 + |U^n(2) - T^n(2)|^2 \Big]. \tag{5.2}$$

Consequently,

$$\mathbb{E}\sum_{i=1}^{n} W^2\big(Q_i(\cdot|U^{i-1}(1),U_i^n(2)),Q_i(\cdot|T^{i-1}(1),T_i^n(2))\big)$$
$$\leq 1/2\cdot(1-\delta)^2\cdot\mathbb{E}\Big[|U^n(1)-T^n(1)|^2+|U^n(2)-T^n(2)|^2\Big]. \tag{5.3}$$

For $U^n(1)=T^n(1)$ *the right-hand-side can be divided by* 2*.*

Proof.
(i): We have

$$\sum_{i=1}^{n}\left|\partial_i\log\frac{Q_i(\zeta_i|u^{i-1}(1),u_i^n(2))}{Q_i(\zeta_i|t^{i-1}(1),t_i^n(2))}\right|^2$$
$$\leq 2\cdot\sum_{i=1}^{n}\left|\partial_i\log Q_i(\zeta_i|u^{i-1}(1),u_i^n(2))-\partial_i\log Q_i(\zeta_i|t^{i-1}(1),u_i^n(2))\right|^2$$
$$+2\cdot\sum_{i=1}^{n}\left|\partial_i\log Q_i(\zeta_i|t^{i-1}(1),u_i^n(2))-\partial_i\log Q_i(\zeta_i|t^{i-1}(1),t_i^n(2))\right|^2. \tag{5.4}$$

To estimate the first term, fix the vectors ζ^n and $u^n(2)$, and consider the mapping

$$g=g(y^n):\mathbb{R}^n\mapsto\mathbb{R}^n,$$

defined by

$$g_i(y^n)=\partial_i\log Q_i(\zeta_i|y^{i-1},u_i^n(2)),\qquad i=1,2,\ldots,n.$$

The first sum on the right-hand-side of (5.4) is just the squared Euclidean norm of the increment of g between points $t^n(1)$, $u^n(1)$. By the contractivity condition for partial derivatives, ((C) of Section 1), the norm of the Jacobian of g is bounded by $1/2\cdot\rho\cdot(1-\delta)$, so the first sum in (5.4) can be bounded by

$$1/4\cdot\rho^2\cdot(1-\delta)^2\cdot|u^n(1)-t^n(1)|^2.$$

A similar bound holds for the second term. Inequality (5.1) is proved. The second statement of (i) is obvious.

(ii): Since the conditional density functions $Q_i(\cdot|\bar{x}_i)$ satisfy the logarithmic Sobolev inequality with constant ρ, we have

$$\sum_{i=1}^{n} D(Q_i(\cdot|U^{i-1}(1),U_i^n(2))\|Q_i(\cdot|T^{i-1}(1),T_i^n(2)))$$
$$\leq\sum_{i=1}^{n}\frac{1}{2\rho}\cdot\mathbb{E}\int_{\mathbb{R}}Q_i(\zeta_i|U^{i-1}(1),U_i^n(2))\left|\partial_i\log\frac{Q_i(\zeta_i|U^{i-1}(1),U_i^n(2))}{Q_i(\zeta_i|T^{i-1}(1),T_i^n(2))}\right|^2 d\zeta_i$$
$$=\frac{1}{2\rho}\cdot\mathbb{E}\int_{\mathbb{R}^n}\sum_{i=1}^{n}\left|\partial_i\log\frac{Q_i(\zeta_i|U^{i-1}(1),U_i^n(2))}{Q_i(\zeta_i|T^{i-1}(1),T_i^n(2))}\right|^2\prod_{i=1}^{n}Q_i(\zeta_i|U^{i-1}(1),U_i^n(2))d\zeta^n. \tag{5.5}$$

Since the product in (5.5) is a density function, (5.1) implies (5.2). To get (5.3) from (5.2), apply the Otto-Villani theorem to each $Q_i(\cdot|t^{i-1}(1), t_i^n(2))$. $\square$

Proof of the Corollary to the Theorem.
Assume that the Theorem holds for $n-1$. Fix i, x^i and $\hat{x}^i$. By the Otto-Villani theorem

$$W^2\big(q_i^n(\cdot|x^i), q_i^n(\cdot|\hat{x}^i)\big) \le \frac{4}{\rho\delta} \cdot D(q_i^n(\cdot|x^i) \| q_i^n(\cdot|\hat{x}^i))$$
$$\le \left(\frac{2}{\rho\delta}\right)^2 \cdot I\big(q_i^n(\cdot|x^i) \| q_i^n(\cdot|\hat{x}^i)\big)$$
$$= \left(\frac{2}{\rho\delta}\right)^2 \cdot \int_{\mathbb{R}^{n-i}} \sum_{k=i+1}^{n} \big| \, \partial_k(\log q_i^n(\zeta_i^n|x^i) - \partial_k(\log q_i^n(\zeta_i^n|\hat{x}^i) \, \big|^2 \, q_i^n(\zeta_i^n|x^n) d\zeta_i^n.$$

Now we use the argument of the proof of Lemma 3. Fix ζ_i^n, and consider the mapping

$$g : \mathbb{R}^i \mapsto \mathbb{R}^{n-i},$$

defined as

$$g_k(x^i) = \partial_k \log q_i^n(\zeta_i^n|x^i), \qquad i < k \le n.$$

The sum within the last integral is the increment of g between points x^i and $\hat{x}^i$. The Jacobian of $g(x^i)$ is the matrix

$$\Big(\partial_{j,k} \log q^n(x^i, \zeta_i^n)\Big), \quad (1 \le j \le i, i < k \le n),$$

i.e., a minor of $-B_1(x^n, x^n, \zeta^n)$. Therefore, by the Contractivity Condition, the norm of the Jacobian of $g(x^i)$ is at most $1/2 \cdot \rho \cdot (1-\delta)$. Thus

$$\sum_{k=i+1}^{n} \big| \, \partial_k \log q_i^n(\zeta_i^n|x^i) - \partial_k \log q_i^n(\zeta_i^n|\hat{x}^i) \, \big|^2 \le 1/4 \cdot (1-\delta)^2 \cdot \rho^2 \cdot |x^i - \hat{x}^i|^2,$$

and (1.4) follows. $\square$

We shall need the Markov operator defined by the system of conditional density functions $\mathcal{Q}$ as follows.

Definition.
Let G be a Markov operator on the probability measures π^n (on $\mathbb{R}^n$) whose transition function is

$$G(v^n|u^n) = \prod_{i=1}^{n} Q_i(v_i|v^{i-1}, u_i^n).$$

Note that if a density function q^n has conditional density functions Q_i then it is invariant with respect to G.

Lemma 4.
(i) If dist(U^n) and dist(V^n) have a joining dist(U^n, V^n) such that

$$\text{dist}(V_i|V^{i-1}, U_i^n) = Q_i(\cdot|V^{i-1}, U_i^n), \quad 1 \le i \le n, \tag{5.6}$$

then dist(V^n) = dist(U^n)G.
(ii) If the system of conditional density functions $\mathcal{Q}$ satisfies condition $LSI\&C(\rho, 1/2(1-\delta))$, then for any two distributions π^n and σ^n on $\mathbb{R}^n$,

$$W^2(\sigma^n G, \pi^n G) \le (1-\delta)^2 W^2(\sigma^n, \pi^n),$$

i.e., G is a strict contraction with respect to W-distance. Consequently, for G there exists a unique invariant distribution q^n.

Proof.
(i) We show that $\text{dist}(U^n)$ and relations (5.6) uniquely determine $\text{dist}(V^n)$. Indeed, it is clear that $\text{dist}(U_1^n, V_1)$ is well-defined by $\text{dist}(U^n)$ and (5.6). Then we see that $\text{dist}(U_2^n, V_1, V_2)$ is well-defined, too, and, by recursion, $\text{dist}(U_i^n, V^i)$ is well-defined for all i. For $i = n$ we get that $\text{dist}(V^n)$ is well-defined.

(ii) Write

$$\sigma^n = \text{dist}\big(U^n(2)\big), \qquad \sigma^n G = \text{dist}\big(U^n(1)\big)$$
$$\pi^n = \text{dist}\big(T^n(2)\big), \qquad \pi^n G = \text{dist}\big(T^n(1)\big).$$

Let us be given a joint distribution for $(U^n(2), T^n(2))$. We inductively build up a joint distribution for the quadruple $(U^n(1), T^n(1), U^n(2), T^n(2))$ as follows. Let us assume that, for some i, $1 \le i \le n$, we have already defined

$$\text{dist}\big(U^{i-1}(1), T^{i-1}(1), U^n(2), T^n(2)\big).$$

Then put

$$\begin{aligned}
&\text{dist}\big(U_i(1)|U^{i-1}(1), T^{i-1}(1), U^n(2), T^n(2)\big) = Q_i\big(\cdot|U^{i-1}(1), U_i^n(2)\big),\\
&\text{dist}\big(T_i(1)|U^{i-1}(1), T^{i-1}(1), U^n(2), T^n(2)\big) = Q_i\big(\cdot|T^{i-1}(1), T_i^n(2)\big),
\end{aligned} \tag{5.7}$$

and define

$$\text{dist}\big(U_i(1), T_i(1)|U^{i-1}(1), T^{i-1}(1), U^n(2), T^n(2)\big)$$

as that joining of the conditional distributions (5.7) that achieves W-distance for every value of the conditions. By (5.3) of Lemma 3,

$$\mathbb{E}|U^n(1) - T^n(1)|^2 = \sum_{i=1}^n \mathbb{E} W^2\Big(Q_i\big(\cdot|U^{i-1}(1), U_i^n(2)\big), Q_i\big(\cdot|T^{i-1}(1), T_i^n(2)\big)\Big)$$
$$\le 1/2 \cdot (1-\delta)^2 \cdot \mathbb{E}\Big[|U^n(1) - T^n(1)|^2 + |U^n(2) - T^n(2)|^2\Big].$$

Rearranging terms we get

$$\mathbb{E}\big|U^n(1) - T^n(1)\big|^2 \le \frac{1/2 \cdot (1-\delta)^2}{1 - 1/2 \cdot (1-\delta)^2} \cdot \big|U^n(2) - T^n(2)\big|^2$$
$$\le (1-\delta)^2 \cdot \big|U^n(2) - T^n(2)\big|^2. \qquad \square$$

6. Estimating the sum in (2.3).

The sum in the right-hand-side of (2.3) is $\sum_{t=1}^{\infty} D_t$, where

$$D_t = \sum_{i=1}^{n} D\big(Y_i(t)|Y^{i-1}(t), Y_i^n(t+1)\|Q_i(\cdot|Y^{i-1}(t), Y_i^n(t+1))\big).$$

In this section we estimate $\sum_{t=1}^{\infty} D_t$ in terms of D_0 and D_1.

Lemma 5. *Assume that the conditional density functions $Q_i(\cdot|\bar{x}_i)$ satisfy condition $LSI\&C(\rho, 1/2(1-\delta))$. Then for the the process (4.2)*

$$\sum_{t=1}^{\infty} D_t \le \frac{1/2 \cdot (1-\delta)^2 \cdot D_0 + D_1}{1-(1-\delta)^2}. \tag{6.1}$$

Proof. It is enough to prove the recursive bound

$$D_t \le 1/2 \cdot (1-\delta)^2 \cdot (D_{t-2} + D_{t-1}), \quad (t \ge 2), \tag{6.2}$$

since (6.2) will imply (6.1) by a simple calculation.

Let us estimate the i'th term in the definition of D_t. We claim that for $t \ge 2$

$$\begin{aligned} &D\big(Y_i(t)|Y^{i-1}(t), Y_i^n(t+1)\|Q_i(\cdot|Y^{i-1}(t), Y_i^n(t+1))\big) \\ &\le D\big(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1))\|Q_i(\cdot|\eta^{i-1}(t), \eta_i^n(t+1))\big). \end{aligned} \tag{6.3}$$

To prove (6.3), recall that, by Lemma 1 and property (v) of the process $(Y^n(t), \eta^n(t))$, the joint distribution $\mathrm{dist}(Y^n(t), Y^n(t+1))$ is the loosely connected copy of $\mathrm{dist}(\eta^n(t), \eta^n(t+1))$, thus

$$\begin{aligned} &D\big(Y_i(t)|Y^{i-1}(t), Y_i^n(t+1)\|Q_i(\cdot|Y^{i-1}(t), Y_i^n(t+1))\big) \\ &= D\big(\eta_i(t)|\eta^{i-1}(t), \eta_i^n(t+1)\|Q_i(\cdot|\eta^{i-1}(t), \eta_i^n(t+1))\big). \end{aligned} \tag{6.4}$$

By Lemma 2 the right-hand-side can be bounded as follows:

$$\le D\big(\eta_i(t)|\eta^{i-1}(t), \eta_i^n(t+1), Y^{i-1}(t-2), Y_i^n(t-1)\|Q_i(\cdot|\eta^{i-1}(t), \eta_i^n(t+1))\big). \tag{6.5}$$

To handle (6.5), we need the following relation of conditional independence:

$$\eta_i(t) \to (Y^{i-1}(t-2), Y_i^n(t-1)) \to (\eta^{i-1}(t), \eta_i^n(t+1)). \tag{6.6}$$

Assume for a moment that (6.6) is proved. Since by definition

$$\mathrm{dist}(\eta_i(t)|(Y^{i-1}(t-2), Y_i^n(t-1)) = Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1)), \tag{6.7}$$

(6.6) implies that the right-hand-side of (6.5) is just

$$D\big(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1))\|Q_i(\cdot|\eta^{i-1}(t), \eta_i^n(t+1))\big),$$

and (6.3) follows.

So let us prove (6.6). By construction $\eta_i(t)$ is conditionally independent of all the process $(Y^n(s), \eta^n(s))$ given $(Y^i(t-2), Y_i^n(t-1))$. (C.f. property (iv) in Section 4 and the definition of $\mathcal{Q}$-extension in Section 3.) Thus (6.6) will be proved as soon as we show that

$$Y_i(t-2) \to (Y^{i-1}(t-2), Y_i^n(t-1)) \to (\eta^{i-1}(t), \eta_i^n(t+1)). \tag{6.8}$$

We know that $(Y^n(t-2), Y^n(t-1))$ is loosely connected, i.e.,

$$Y_i(t-2) \to (Y^{i-1}(t-2), Y_i^n(t-1)) \to Y^i(t-1).$$

So for the proof of (6.8) it is enough to show that

$$Y_i(t-2) \to (Y^{i-1}(t-2), Y^n(t-1)) \to (\eta^{i-1}(t), \eta_i^n(t+1)). \tag{6.9}$$

This follows from the relations

$$\begin{aligned}&\eta^{i-1}(t) \quad \text{is conditionally independent of}\\ &\text{everything, given} \quad (Y^{i-1}(t-2), Y^n(t-1)),\end{aligned}$$

$$\begin{aligned}&\eta^n(t+1) \quad \text{is conditionally independent of}\\ &\text{everything, given} \quad (Y^n(t-1), Y^n(t)),\end{aligned}$$

and

$$Y^n(t-2) \to Y^n(t-1) \to Y^n(t).$$

Thus (6.8), and consequenyly (6.6) and (6.3) are proved.

By (6.3) and Lemma 3 we have for $t \geq 2$

$$\begin{aligned} D_t &\leq \sum_{i=1}^n D\big(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1)) \| Q_i(\cdot|\eta^{i-1}(t), \eta_i^n(t+1))\big) \\ &\leq 1/4 \cdot (1-\delta)^2 \cdot \rho \cdot \mathbb{E}\Big[|Y^n(t-2) - \eta^n(t)|^2 + |Y^n(t-1) - \eta^n(t+1)|^2\Big]. \end{aligned} \tag{6.10}$$

To bound the right-hand-side of (6.10) by D_{t-2} and D_{t-1}, recall that

$$\begin{aligned} &\mathbb{E}\big\{|Y_i(t-2) - \eta_i(t)|^2 \;\big|\; Y^{i-1}(t-2), Y_i^n(t-1)\big\} \\ &= \mathbb{E}W^2\Big(\operatorname{dist}(Y_i(t-2)|Y^{i-1}(t-2), Y_i^n(t-1)), Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t))\Big). \end{aligned}$$

Therefore, by the Otto-Villani theorem,

$$\mathbb{E}\big|Y^n(t-2)-\eta^n(t)\big|^2 = \sum_{i=1}^n \mathbb{E}\Big\{\mathbb{E}\big\{|Y_i(t-2)-\eta_i(t)|^2 \;\big|\; Y^{i-1}(t-2), Y_i^n(t-1)\big\}\Big\}$$
$$\leq \frac{2}{\rho}\cdot\sum_{i=1}^n D\big(Y_i(t-2)|Y^{i-1}(t-2), Y_i^n(t-1)\|Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1))\big)$$
$$= \frac{2}{\rho}\cdot D_{t-2}. \tag{6.11}$$

Inequalities (6.10-6.11) imply (6.2) and hence (6.1). □

Remark. By the definition of W-distance and by (6.11)

$$W^2(Y^n(t), Y^n(t+2)) \leq \mathbb{E}|Y^n(t-2)-\eta^n(t)|^2 \leq \frac{2}{\rho}\cdot D_{t-2},$$

thus

$$W^2(Y^n(t-2), Y^n(t)) \to 0 \quad \text{exponentially fast.} \tag{6.12}$$

We shall need this to prove that $(Y^n(2t-1), Y^n(2t))$ converges, along some subsequence, in W-distance.

7. Convergence properties and proof of the Main Lemma

Before we proceed to the proof of the Main Lemma , we establish Wasserstein-convergence for the process $\big(Y^n(t)\big)$.

Lemma 6. *If the conditional density functions $Q_i(\cdot|\bar{x}_i)$ satisfy condition $LSI\&C(\rho, 1/2(1-\delta))$ then there exists a joint distribution dist(X^n, Z^n) such that, along some subsequence of the integers t,*

$$\lim_{t\to\infty} (Y^n(t), Y^n(t+1)) = (X^n, Z^n) \quad in \quad W\text{-}distance. \tag{7.1}$$

Moreover,

$$dist(X^i, Z_i^n) = q^n \quad for\ all \quad i. \tag{7.2}$$

Proof. We claim that there exists a joint distribution $\mathrm{dist}(X^n, Z^n, T^n)$ such that

$$\begin{aligned}&\big(Y^n(t-1), \eta^n(t+1), Y^n(t), \eta^n(t+2), Y^n(t+1), \eta^n(t+3)\big)\\ &\to \big(X^n, X^n, Z^n, Z^n, T^n, T^n\big) \quad \text{in} \quad W\text{-distance},\end{aligned} \tag{7.3}$$

along some subsequence of the values of t. Indeed, the sequences $(Y^n(2t))$ and $(Y^n(2t+1))$ being Cauchy with respect to W-distance, the process in (7.1) is uniformly square integrable, hence tight, hence compact with respect to weak convergence, and hence also for W-convergence. Since

$$\lim_{t\to\infty} E|Y^n(t-2) - \eta^n(t)|^2 = 0,$$

(7.3) follows.

To prove (7.2), we deduce from (7.3) the following relations: for all i,

$$\mathrm{dist}(X_i|X^{i-1}, Z_i^n) = Q_i(\cdot|X^{i-1}, Z_i^n), \qquad \mathrm{dist}(Z_i|Z^{i-1}, T_i^n) = Q_i(\cdot|Z^{i-1}, T_i^n), \tag{7.4}$$

and

$$\mathrm{dist}(X^n) = \mathrm{dist}(T^n). \tag{7.5}$$

The latter is clear, since $\lim_{t\to\infty} W^2(Y^n(t), Y^n(t+2)) = 0$. The first line of (7.4) follows from the relations

$$(Y^{i-1}(t-1), \eta_i(t+1), Y_i^n(t)) \to (X^i, Z_i^n) \quad \text{in distribution},$$

and

$$\mathrm{dist}(\eta_i(t+1)|Y^{i-1}(t-1), Y_i^n(t)) = Q_i(\cdot|Y^{i-1}(t-1), Y_i^n(t)).$$

The second line can be proved similarly.

By Lemma 4, (7.4) implies that

$$\mathrm{dist}(X^n) = \mathrm{dist}(Z^n)G, \quad \text{and} \quad \mathrm{dist}(Z^n) = \mathrm{dist}(T^n)G.$$

By (7.5) this means that $\mathrm{dist}(X^n) = \mathrm{dist}(Z^n)$ is invariant with respect to G, thus by Lemma 4 it must be q^n. This, together with (7.4), imlies (7.2). $\square$

Proof of the Main Lemma.
By the Auxiliary Theorem, it is enough to prove that

$$\lim_{t\to\infty} D(Y^n(t)\|X^n) = 0 \quad \text{along some subsequence.} \tag{7.6}$$

Since $\mathrm{dist}(Y^n(t)) = \mathrm{dist}(\eta^n(t))$, (7.6) is equivalent to

$$\lim_{t\to\infty} D(\eta^n(t)\|X^n) = 0 \quad \text{along some subsequence.}$$

We shall move along a subsequence such that

$$\lim_{t\to\infty} (Y^n(t-2), Y^n(t-1)) = (X^n, Z^n) \quad \text{in} \quad W\text{-distance.}$$

By Lemma 2 and relations (6.6-6.7) we have

$$\begin{aligned} D(\eta^n(t)\|X^n)) &= \sum_{i=1}^n D(\eta_i(t)|\eta^{i-1}(t))\|q_i(\cdot|\eta^{i-1}(t)) \\ &\le \sum_{i=1}^n D(\eta_i(t)|\eta^{i-1}(t), Y^{i-1}(t-2), Y_i^n(t-1))\|q_i(\cdot|\eta^{i-1}(t)) \\ &= \sum_{i=1}^n D(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1))\|q_i(\cdot|\eta^{i-1}(t)), \end{aligned} \tag{7.7}$$

where $q_i(\cdot|y^{i-1}) = \mathrm{dist}(X_i|X^{i-1} = y^{i-1})$.

We are going to replace $q_i(\cdot|\eta^{i-1}(t))$ by $q_i(\cdot|X^{i-1})$, by the cost of a small error term. This requires a tedious approximation argument which we carry out in the Appendix, proving the following lemma.

Lemma 7. *Let dist$(X^n) = q^n$ be fixed, and assume that its conditional density functions $Q_k(\cdot|\bar{x}_k)$ satisfy condition LSI&C$(\rho, 1/2(1-\delta))$. Fix an $\varepsilon > 0$. There exists a $\gamma > 0$ such that for any i and any joint distribution dist$(\bar{Y}_i, \eta^{i-1}, X^n)$ we have the implication*

$$\begin{aligned} &\left\{ \mathbb{E}|\bar{Y}_i - \bar{X}_i|^2 < \gamma^2 \quad \& \quad \mathbb{E}|\eta^{i-1} - X^{i-1}|^2 < \gamma^2 \right\} \\ &\implies D(Q_i(\cdot|\bar{Y}_i)\|q_i(\cdot|\eta^{i-1})) \le D(Q_i(\cdot|\bar{Y}_i)\|q_i(\cdot|X^{i-1})) + \varepsilon. \end{aligned} \tag{7.8}$$

We give the proof in the Appendix, and apply the lemma now.

Fix an $\varepsilon > 0$, and find the corresponding $\gamma > 0$. Select t so large that

$$W^2((Y^n(t-2), Y^n(t-1)), (X^n, Z^n)) \leq \gamma^2/4,$$

and

$$\mathbb{E}|\eta^n(t) - Y^n(t-2)| \leq \gamma^2/4,$$

where (X^n, Z^n) is defined in Lemma 6. Define the joint distribution $\mathrm{dist}(Y^n(t-2), Y^n(t-1), X^n, Z^n)$ to achieve

$$\mathbb{E}|(Y^n(t-2), Y^n(t-1)) - (X^n, Z^n)|^2| \leq \gamma^2/4;$$

this also ensures $\mathbb{E}|\eta^{i-1}(t) - X^{i-1}|^2 \leq \gamma^2$ for all i.

We apply Lemma 7 with $(Y^{i-1}(t-2), Y_i^n(t-1))$ in place of $\bar{Y}_i$, with $\eta^{i-1}(t)$ in place of η^{i-1} and with (X^i, Z_i^n) in place of X^n. Using also Lemma 2 and (7.2) of Lemma 6:

$$\begin{aligned}
&D(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1)))\|q_i(\cdot|\eta^{i-1})) \\
&\leq D(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1))\|q_i(\cdot|X^{i-1}) + \varepsilon \\
&\leq D(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1))\|Q_i(\cdot|X^{i-1}, Z_i^n) + \varepsilon. \qquad (7.9)
\end{aligned}$$

Now we can apply Lemma 3 with $(Y^n(t-2), Y^n(t-1))$ in place of $(U^n(1), U^n(2))$, and with (X^n, Z^n) in place of $(T^n(1), T^n(2))$. We get

$$\begin{aligned}
&\sum_{i=1}^n D(Q_i(\cdot|Y^{i-1}(t-2), Y_i^n(t-1))\|Q_i(\cdot|X^{i-1}, Z_i^n)) \\
&\leq 1/4(1-\delta)^2 \cdot \rho \cdot \left[|Y^n(t-2) - X^n|^2 + |Y^n(t-1) - Z^n|^2\right] \leq \gamma^2. \qquad (7.10)
\end{aligned}$$

Inequalities (7.7), (7.9) and (7.10) imply

$$D(Y^n(t)\|X^n) \leq D(\eta^n(t)\|X^n) \leq \varepsilon + \gamma^2.$$

The Main Lemma is proved. $\square$

8. Bounds by the Fisher information: end of the proof

To derive the Theorem from the Main Lemma, we specialize the process $(Y^n(t))$, taking for $\mathrm{dist}(Y^n(0), Y^n(1))$ the loosely connected copy of $(Y^n(0), Y^n(0))$, where $\mathrm{dist}(Y^n(0)) = p^n$. Thus, in particular,

$$\mathrm{dist}(Y^i(0), Y_i^n(1)) = p^n \quad 1 \le i \le n.$$

The joint distribution $\mathrm{dist}(Y^n(0), Y^n(1))$ does determine $\mathrm{dist}(Y^n(0), Y^n(1), \eta^n(2))$, but we are free to select $\mathrm{dist}(Y^n(1), Y^n(2))$ subject to the conditions that it be loosely connected, and $\mathrm{dist}(Y^n(2)) = \mathrm{dist}(\eta^n(2))$. We define $\mathrm{dist}(Y^n(1), Y^n(2))$ as follows. Let us extend $\mathrm{dist}(Y^n(0), Y^n(1), \eta^n(2))$ to a joint distribution $\mathrm{dist}(\eta^n(1), Y^n(0), Y^n(1), \eta^n(2))$ in such a way that $\mathrm{dist}(\eta^n(1), Y^n(0))$, too, be the loosely connected copy of $(Y^n(0), Y^n(0))$ as well, and the following Markov relations hold:

$$\eta^n(1) \to Y^n(0) \to Y^n(1), \qquad \eta^n(1) \to (Y^n(0), Y^n(1)) \to \eta^n(2). \tag{8.1}$$

Then let $\mathrm{dist}(Y^n(1), Y^n(2))$ be the loosely connected copy of $(\eta^n(1), \eta^n(2))$.

For this case we can estimate D_0 and D_1 in terms of $I(p^n \| q^n)$. For D_0 this is easy. Since $\mathrm{dist}(Y^i(0), Y_i^n(1)) = p^n$ for all i, we have

$$\begin{aligned} D_0 &= \sum_{i=1}^n D\big(Y_i(0) | Y^{i-1}(0), Y_i^n(1) \| Q_i(\cdot | Y^{i-1}(0), Y_i^n(1))\big) \\ &= \sum_{i=1}^n D\big(Y_i(0) | \bar{Y}_i(0) \| Q_i(\cdot | \bar{Y}_i(0))\big). \end{aligned}$$

Applying the logarithmic Sobolev inequality for $Q_i(\cdot | \bar{Y}_i(0))$, we get that

$$D_0 \le \frac{1}{2\rho} I(p^n \| q^n). \tag{8.2}$$

The estimate for D_1 is more complicated. By the definition of $\mathrm{dist}(Y^n(1), Y^n(2))$:

$$\begin{aligned} D_1 &= \sum_{i=1}^n D\big(Y_i(1) | Y^{i-1}(1), Y_i^n(2) \| Q_i(\cdot | Y^{i-1}(1), Y_i^n(2))\big) \\ &= \sum_{i=1}^n D\big(\eta_i(1) | \eta^{i-1}(1), \eta_i^n(2) \| Q_i(\cdot | \eta^{i-1}(1), \eta_i^n(2))\big). \end{aligned} \tag{8.3}$$

To overbound the terms in the last line of (8.3) we need the Markov relation

$$\eta_i(1) \to \big(\eta^{i-1}(1), Y_i^n(0)\big) \to \eta_i^n(2). \tag{8.4}$$

By the loose connection of $(\eta^n(1), Y^n(0))$, for (8.4) it is enough to prove that

$$\eta_i(1) \to \big(\eta^{i-1}(1), Y^n(0)\big) \to \eta_i^n(2)$$

or

$$\eta^n(1) \to Y^n(0) \to \eta^n(2).$$

This follows from the definition of $\eta^n(1)$ (see (8.1)).

Using (8.4), Lemma 3 and the equality $\operatorname{dist}(\eta^i(1), Y_i^n(0)) = p^n$ we get

$$\begin{aligned}
&D\big(\eta_i(1)|\eta^{i-1}(1), \eta_i^n(2)\|Q_i(\cdot|\eta^{i-1}(1), \eta_i^n(2))\big)\\
&\le D\big(\eta_i(1)|\eta^{i-1}(1), Y_i^n(0)\|Q_i(\cdot|\eta^{i-1}(1), \eta_i^n(2))\big)\\
&= D\big(p_i(\cdot|\eta^{i-1}(1), Y_i^n(0))\|Q_i(\cdot|\eta^{i-1}(1), \eta_i^n(2))\big).
\end{aligned}$$

By the logarithmic Sobolev inequality, this implies

$$\begin{aligned}
&D\big(\eta_i(1)|\eta^{i-1}(1), \eta_i^n(2)\|Q_i(\cdot|\eta^{i-1}(1), \eta_i^n(2))\big)\\
&\le \frac{1}{2\rho}\cdot \mathbb{E} I\big(p_i(\cdot|\eta^{i-1}(1), Y_i^n(0))\|Q_i(\cdot|\eta^{i-1}(1), \eta_i^n(2))\big).
\end{aligned}$$

Substituting this into (8.3) we get

$$D_1 \le \frac{1}{2\rho}\cdot \sum_{i=1}^n \mathbb{E}\int_{\mathbb{R}} p_i(y_i|\eta^{i-1}(1), Y_i^n(0))\left|\partial_i \log \frac{p_i(y_i|\eta^{i-1}(1), Y_i^n(0))}{Q_i(y_i|\eta^{i-1}(1), \eta_i^n)(2)}\right|^2 dy_i.$$

Using the relation $\operatorname{dist}(\eta^{i-1}(1), Y_i^n(0)) = \bar{p}_i$ and the bound $(a+b)^2 \le 2a^2 + 2b^2$, the last line can be continued as follows:

$$\begin{aligned}
D_1 &\le 2\cdot\frac{1}{2\rho}\cdot \sum_{i=1}^n \mathbb{E}\int_{\mathbb{R}} p_i(y_i|\bar{y}_i)\left|\partial_i \log \frac{p_i(y_i|\bar{y}_i)}{Q_i(y_i|\bar{y}_i)}\right|^2 dy_i\\
&+ 2\cdot\frac{1}{2\rho}\cdot \sum_{i=1}^n \mathbb{E}\int_{\mathbb{R}} p_i(y_i|\eta^{i-1}(1), Y_i^n(0))\left|\partial_i \log \frac{Q_i(y_i|\eta^{i-1}(1), Y_i^n(0))}{Q_i(y_i|\eta^{i-1}(1), \eta_i^n(2))}\right|^2 dy_i.
\end{aligned} \tag{8.5}$$

The first term on the right-hand-side of (8.5) is $\frac{1}{\rho}I(p^n\|q^n)$. The second term can be written as

$$\frac{1}{\rho}\cdot \mathbb{E}\int_{\mathbb{R}^n} \sum_{i=1}^n \left|\partial_i \log \frac{Q_i(u_i|\eta^{i-1}(1), Y_i^n(0))}{Q_i(u_i|\eta^{i-1}(1), \eta_i^n(2))}\right|^2 \prod_{i=1}^n p_i(u_i|\eta^{i-1}(1), Y_i^n(0))du^n.$$

By Lemma 3 for any $u^n \in \mathbb{R}^n$ we have the inequality

$$\sum_{i=1}^n \left|\partial_i \log \frac{Q_i(u_i|\eta^{i-1}(1), Y_i^n(0))}{Q_i(u_i|\eta^{i-1}(1), \eta_i^n(2))}\right|^2 \le 1/4\cdot\rho\cdot(1-\delta)^2\cdot|Y^n(0) - \eta^n(2)|^2,$$

so the second term in the right-hand-side of (8.5) is bounded by

$$1/4 \cdot \rho \cdot (1-\delta)^2 \cdot \mathbb{E}|Y^n(0) - \eta^n(2)|^2. \tag{8.6}$$

By the definition of $\mathrm{dist}(Y^n(0), Y^n(1), \eta^n(2))$ and by the Otto-Villani theorem (8.6) can be continued as follows:

$$\begin{aligned}
&\le 1/4 \cdot \rho \cdot (1-\delta)^2 \cdot \sum_{i=1}^{n} \mathbb{E}\Big\{ \big\{ \mathbb{E}|Y_i(0) - \eta_i(2)|^2 | Y^{i-1}(0), Y_i^n(1) \big\} \Big\} \\
&\le 1/2 \cdot (1-\delta)^2 \cdot \sum_{i=1}^{n} D\big(Y_i(0)|Y^{i-1}(0), Y_i^n(1) \| Q_i(\cdot|Y^{i-1}(0), Y_i^n(1))\big) \\
&\le 1/4 \cdot (1-\delta)^2 \cdot \frac{1}{\rho} \cdot I(p^n \| q^n). \qquad (8.7)
\end{aligned}$$

From (8.5-8.7) we get the estimate

$$D_1 \le \frac{1}{\rho}\big[1 + 1/4(1-\delta)^2\big] \cdot I(p^n \| q^n).$$

Since $D_0 \le \frac{1}{\rho} \cdot I(p^n \| q^n)$, the Main Lemma yields

$$D(p^n \| q^n) \le \frac{1}{\rho\delta} \cdot I(p^n \| q^n) \cdot \frac{1 + 1/2(1-\delta)^2}{(2-\delta)} \le \frac{1}{\rho\delta} \cdot I(p^n \| q^n).$$

This completes the proof of the Theorem. $\square$

Appendix

Proof of Lemma 7.
For the joint distribution $\operatorname{dist}(\bar{Y}_i, \eta^{i-1}, X^n)$ consider the conditonal relative entropy

$$D(Q_i(\cdot|\bar{Y}_i)\|q_i(\cdot|\eta^{i-1})) = \int_{\mathbb{R}^{n-1}\times\mathbb{R}^{i-1}} D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1}))d\pi(\bar{y}_i, z^{i-1}), \tag{A.1}$$

where integration is with respect to $\operatorname{dist}(\bar{Y}_i, \eta^{i-1})$. Our first aim is to restrict integration in (A.1) to a smaller set, by the cost of a small error term. This smaller set will be compact, and will have other useful properties.

First we prove, for fixed $\bar{y}_i$ and z^{i-1}, the following estimate which allows to control the truncation by L_2-techniques:

$$\begin{aligned} &D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1})) \\ &\leq \rho^2 \cdot (1-\delta)^2 \cdot \left[|y^{i-1} - z^{i-1}|^2 + \int_{\mathbb{R}^{n-i}} |y_i^n - s_i^n|^2 \cdot q_i^n(s_i^n|z^{i-1})ds_i^n\right]. \end{aligned} \tag{A.2}$$

To prove (A.2), let us introduce an auxiliary random variable S_i^n, putting

$$\operatorname{dist}(S_i^n|\eta^{i-1}, \bar{Y}_i) = q_i^n(\cdot|\eta^{i-1}), \tag{A.3}$$

where $q_i^n(\cdot|x^{i-1}) = \operatorname{dist}(X_i^n|X^{i-1} = x^{i-1})$. Let $S_i^n \to (\bar{Y}_i, \eta_{i-1}) \to X^n$. Thereby we have defined $\operatorname{dist}(X^n, \bar{Y}_i, \eta^{i-1}, S_i^n)$.

By Lemma 2 we have

$$D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1})) \leq D(Q_i(\cdot|\bar{y}_i)\|Q_i(\cdot|z^{i-1}, S_i^n)), \tag{A.4}$$

where $\operatorname{dist}(S_i^n) = q_i^n(\cdot|z^{i-1})$.

By the logarithmic Sobolev inequality for $Q_i(\cdot|z^{i-1}, S_i^n)$, the right-hand-side of (A.4) can be estimated as follows:

$$\begin{aligned} D(Q_i(\cdot|\bar{y}_i)\|Q_i(\cdot|z^{i-1}, S_i^n)) &\leq \frac{1}{2\rho}\int_{\mathbb{R}^{n-i}} I\big(Q_i(\cdot|\bar{y}_i)\|Q_i(\cdot|z^{i-1}, s_i^n)\big) \cdot q_i^n(s_i^n|z^{i-1})ds_i^n \\ = \frac{1}{2\rho}\int_{\mathbb{R}^{n-i}} &\left[\int Q_i(\eta|\bar{y}_i)|\partial_i V(y^{i-1}, \eta, y_i^n) - \partial_i V(z^{i-1}, \eta, s_i^n)|^2 d\eta\right] q_i^n(s_i^n|z^{i-1})ds_i^n. \end{aligned} \tag{A.5}$$

By Lemma 3 we have the following bound for the inner integral:

$$\begin{aligned} &\int_{\mathbb{R}} Q_i(\eta|\bar{y}_i)\ \big|\ \partial_i V(y^{i-1}, \eta, y_i^n) - \partial_i V(z^{i-1}, \eta, s_i^n)\ \big|^2\ d\eta \\ &\leq \rho^2 \cdot (1-\delta)^2 \cdot \left[|y^{i-1} - z^{i-1}|^2 + |y_i^n - s_i^n|^2\right]. \end{aligned} \tag{A.6}$$

Estimates (A.5-A.6) imply (A.2).

To exploit (A.2), we need bounds for the integrals of $|\bar{Y}_i|^2$, $|\eta^{i-1}|^2$ and $|S_i^n|^2$ over sets of small probability.

Assume

$$\mathbb{E}|\eta^{i-1} - X^{i-1}|^2 < \varepsilon^2 \quad \text{and} \quad \mathbb{E}|\bar{Y}_i - \bar{X}_i|^2 < \varepsilon^2.$$

Then for any set E, measurable with respect to $(X^n, \bar{Y}_i, \eta^{i-1}, S_i^n)$, we have

$$\begin{aligned} &\mathbb{E}\big\{|\bar{Y}_i|^2 \cdot \chi_E\big\} \leq 2 \cdot \mathbb{E}\big\{|\bar{X}_i|^2 \cdot \chi_E\big\} + 2\varepsilon^2, \\ &\mathbb{E}\big\{|\bar{\eta}^{i-1}|^2 \cdot \chi_E\big\} \leq 2 \cdot \mathbb{E}\big\{|X^{i-1}|^2 \cdot \chi_E\big\} + 2\varepsilon^2. \end{aligned} \tag{A.7}$$

To handle $|S_i^n|^2$, we recall the Corollary to the Theorem (1.2), saying that

$$W^2(S_i^n, X_i^n) \leq C^2 \cdot W^2(\eta^{i-1}, X^{i-1}) \leq C^2 \cdot \varepsilon^2. \tag{A.8}$$

where C is a constant depending on δ. (Recall that for this we only need the Theorem for $n-1$.) By (A.8),

$$\mathbb{E}\big\{|S_i^n|^2 \cdot \chi_E\big\} \leq 2 \cdot \mathbb{E}\big\{|X_i^n|^2 \cdot \chi_E\big\} + 2C^2 \cdot \varepsilon^2. \tag{A.9}$$

Let $K \subset \mathbb{R}$ be a compact interval such that the integral of $|X^n|^2$ outside of K^n is less than ε^2. Then, by (A.2), (A.7) and (A.9)

$$\begin{aligned} &D(Q_i(\cdot|\bar{Y}_i)\|q_i(\cdot|\eta^{i-1})) \\ &\leq \int_{\mathbb{K}^{n-1}\times\mathbb{K}^{i-1}} D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1}))d\pi(\bar{y}_i, z^{i-1}) + O(\varepsilon^2), \end{aligned} \tag{A.10}$$

where $O(\varepsilon^2)$ is bounded by an absolute constant times $C^2 \cdot \varepsilon^2$.

Further select $\gamma > 0$ so small that for any m and any measure μ on $\mathbb{R}^n \times \mathbb{R}^m$ whose marginal for the first n coordinates is q^n, we have

$$E \subset \mathbb{R}^n \times \mathbb{R}^m, \quad \mu(E) \leq \gamma \implies \mathbb{E}\big\{|X^n|^2 \cdot \chi_E\big\} < \varepsilon^2.$$

Assume

$$\mathbb{E}|\eta^{i-1} - X^{i-1}|^2 < \gamma^2,$$

and define

$$B = \Big\{(z^{i-1}, x^{i-1}) \in \mathbb{R}^{i-1} \times \mathbb{R}^{i-1} : |z^{i-1} - x^{i-1}| \leq \gamma\Big\}.$$

We have then

$$Pr\big\{(\eta^{i-1}, X^{i-1}) \notin B\big\} \leq \gamma. \tag{A.11}$$

Let $M = (K^{n-1} \times K^{i-1} \times K^{i-1}) \cap B$. Then (A.10) can be continued as follows:

$$D(Q_i(\cdot|\bar{Y}_i)\|q_i(\cdot|\eta^{i-1})) \leq \int_M D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1}))d\mu(\bar{y}_i, z^{i-1}, x^{i-1}) + O(\varepsilon^2), \tag{A.12}$$

where $\mu = \text{dist}(\bar{Y}_i, \eta^{i-1}, X^{i-1})$. (The integrand does not depend on x^{i-1}, but the domain of integration does.) Thus in estimating $D(Q_i(\cdot|\bar{Y}_i\|q_i(\cdot|\eta^{i-1}))$, integration can be restricted to those values of $(\bar{y}_i, z^{i-1}, x^{i-1})$, for which all coordinates lie in K, and $|z^{i-1} - x^{i-1}| < \gamma$.

To overbound the integral (A.12) we are going to use the following easy-to-prove

Facts.
Let $f(x, y) : \mathbb{R}^i \times \mathbb{R}^k \mapsto \mathbb{R}$ be a continuous function such that $f(x, \cdot)$ is integrable for every x. Define

$$g_L(x) = \int_L f(x, y)dy, \quad L \subset \mathbb{R}^k.$$

Then:
(i) For every compact $L \subset \mathbb{R}^k$ $g_L(x)$ is continuous;
(ii) for every compact $M \subset \mathbb{R}^i$ there exists a compact $L \subset \mathbb{R}^k$ such that

$$|g_{\mathbb{R}^k}(x) - g_L(x) \leq \varepsilon, \quad x \in M.$$

It easily follows from these Facts that $q^i(x^i) = \int_{\mathbb{R}^{n-i}} q^n(x^n)dx_i^n$ is continuous in x^i, and therefore so is $q_i(x_i|x^{i-1})$.

Consider the integrand in (A.12):

$$D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1})) = \int_{\mathbb{R}} Q_i(\eta|\bar{y}_i) \log \frac{Q_i(\eta|\bar{y}_i)}{q_i(\eta|z^{i-1})} d\eta. \tag{A.13}$$

By Fact (ii), for $(\bar{y}_i, z^{i-1}) \in K^{n-1} \times K^{i-1}$ we can replace the domain of integration $\mathbb{R}$ by $[-A, A]$, with some finite A, depending on K:

$$\left| D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1})) - \int_{-A}^{A} Q_i(\eta|\bar{y}_i) \log \frac{Q_i(\eta|\bar{y}_i)}{q_i(\eta|z^{i-1})} d\eta \right| < \varepsilon \quad \text{on} \quad K^{n-1} \times K^{i-1}, \tag{A.14}$$

for A large enough. By Fact (i), the integral in (A.14) is uniformly continuous on $K^{n-1} \times K^{i-1}$.

Make on $\gamma > 0$ the additional constraint that the integral in (A.14) changes by less than ε if we replace $z^{i-1} \in K^{i-1}$ by $x^{i-1} \in K^{i-1}$, provided $|z^{i-1} - x^{i-1}| < \gamma$:

$$\left| \int_{-A}^{A} Q_i(\eta|\bar{y}_i) \log \frac{Q_i(\eta|\bar{y}_i)}{q_i(\eta|z^{i-1})} d\eta - \int_{-A}^{A} Q_i(\eta|\bar{y}_i) \log \frac{Q_i(\eta|\bar{y}_i)}{q_i(\eta|x^{i-1})} \right| < \varepsilon \tag{A.15}$$

for $(y^{n-1}, z^{i-1}, x^{i-1}) \in K^{n-1} \times K^{i-1} \times K^{i-1}$, $|z^{i-1} - x^{i-1}| < \gamma$.

Integating (A.15) over M, and taking into account the definition of the set B and the bound (A.14):

$$\begin{aligned} &\int_M D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|z^{i-1}))d\mu(\bar{y}_i, z^{i-1}, x^{i-1}) \\ &\leq \int_M D(Q_i(\cdot|\bar{y}_i)\|q_i(\cdot|x^{i-1}))d\mu(\bar{y}_i, z^{i-1}, x^{i-1}) + O(\varepsilon). \end{aligned}$$

Substituting this into (A.12) we get

$$D(Q_i(\cdot|\bar{Y}_i)\|q_i(\cdot|\eta^{i-1})) \leq D(Q_i(\cdot|\bar{Y}_i)\|q_i(\cdot|X^{i-1})) + O(\varepsilon). \qquad \square$$

References

[Ba-E] D. Bakry, M. Emery, *Diffusions hypercontractives*, Séminaire de Probabilités, XIX. Lecture Notes in Math. 1123, Springer, 1985, pp. 177-206.

[B-G] S. Bobkov, F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), 1-28.

[B-G-L] S. Bobkov, I. Gentil, M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations*, J. Math. Pures Appl. **80** (2001), 669-696.

[Bod-He] Th. Bodineau, B. Helffer, *On Log-Sobolev inequalities for unbounded spin systems*, Funct. Anal. **166,** (1999), 168-178.

R.L. Dobrushin, *Prescribing a system of random variables by conditional distributions*, Theory of Prob. Appl. **15** (1970), 453-486.

[Ge] I. Gentil, *Inegalités de Sobolev logarithmiques et hypercontractivité en mécanique statisique et en E.D.P., thesis, Université Paul Sabatier* (2001).

[Gr] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), 1061-1083.

[He] B. Helffer, *Remarks on decay of correlation and Witten-Laplacians III. Application to logarithmic Sobolev inequalities*, Ann. Inst. H. Poincaré, **35,** (1999), 483-508.

[Ho-S] R. Holley, D. Stroock, *Logarithmic Sobolev inequalities and stochastic Ising models*, J. Statist. Phys. **16** (1987), 1159-1191.

[L-P-Y] C. Landim, G. Panizo, H. T. Yau, *Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems*, Ann. Inst. H. Poincaré Probab. Statist. **38** (2002), 739-777.

[L] M. Ledoux, *Logarithmic Sobolev inequalities for some matrix models and unbounded spin systems, preprint* (1999).

[M1] K.Marton, *Measure concentration for a class of random processes*, Probab. Theory Relat. Fields, **110,** (1998), 427-439,.

[M2] K. Marton, *Measure concentration for Euclidean distance in the case of dependent random variables,*, Ann. Prob. **32** (2004), 2526-2544.

[O-V] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361-400.

[S-Z1] D. Stroock, B. Zegarlinski, *The logarithmic Sobolev inequality for continuous spin systems on a lattice*, J. Funct. Anal. **104** (1992), 299-326.

[S-Z2] D. Stroock, B. Zegarlinski, *The logarithmic Sobolev inequality for discrete spin systems on a lattice*, Comm. Math. Phys. **149** (1992), 175-193.

[S-Z3] D. Stroock, B. Zegarlinski, *The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition*, Comm. Math. Phys. **144** (1992), 303-323.

[T] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geometric and Functional Analysis, **6,** (1996), 587-600.

[V] C. Villani, *Topics in optimal transportation, Graduate Studies in Mathematics, vol. 58*, American Math. Soc., 2003.

[Y1] N. Yoshida, *The log-Sobolev inequality for weakly coupled lattice fields,*, Probab. Theory Relat. Fields, **115,** (1999), 1-10,.

[Y2] N. Yoshida, *The equivalence of the log-Sobolev inequality and a mixing condition for unbounded spin systems on the lattice, preprint* (1999).

[Z] B. Zegarlinski, *The strong decay to equilibrium in the stochastic dynamics of unbounded spin systems on a lattice,*, Commun. Math. Phys. **175,** (1996), 401-432,.

H-1364 POB 127, Budapest, Hungary
E-mail address: `marton@renyi.hu`